Metadata Workshop III – Interagency Chronlog

Tue 8/14/2012

1. 09:00 - Tim Owen – Welcome
    a. ATRAC rollout
    b. Metadata – term 1st used in 1968
2. 09:05 – Christina – Overview / logistics
3. Intros
    a. Jeff De La Beaujardière
    b. Yonsook Enloe - NASA
    c. Ken McDonald
    d. Glenn Rutledge
    e. Ted Habermann
    f. Nancy Ritchey
    g. Lynda Wayne
    h. Jaci Mize
    i. Don Collins
    j. Luther Lighty – NASA/Echo
    k. Katie Baynes – NASA/Echo
    l. Phil Jones
    m. Valerie Toner
    n. Heather Brown
    o. Ken Roberts
    p. Jason Cooper
    q. Curt Tilmes - GCIS
    r. Ana Pineiro Privette
    s. Linda Copley
    t. Lola Olsen
    u. Scott Ritz
    v. Jay Morris
    w. Tammy Beatty – ORNL
    x. Christina Lief
    y. <<phone>>
    z. Anusha – NGDC
    aa. ?? GCR – NASA
    bb. Shaida Johnson
    cc. Geoff Goodrum
    dd. Ken O'Brien
    ee. ?? NODC (phone – could not hear clearly)
    ff. Leann Cross (?) - USGS Woods Hole
4. Post workshop content
    a. Wiki
    b. GoogleDocs? (TBD)
5. 0922 - Jeff DLB – NOAA Data Mgmt
    a. Vision

i. Discoverable
ii. Accessible
iii. Well-documented
iv. Preserved

b. NOAA Environmental Data Mgmt Framework
   i. Still in draft
   ii. Due Mar 2013
   iii. Based on interagency National Earth Observations Strategy, CH 4
c. Framework
   i. Principles
   ii. Governance
   iii. Resources
   iv. Standards
   v. Architecture
   vi. Assessment
   vii. Data lifecycle
d. Governance
   i. NOAA DM Governance org chart
      1. David Layton – NOAA enterprise architect
   ii. Procedural directives (PD)
      1. Data Mgmt Planning PD
         a. Should be done before project is executed
         b. For expensive projects, part of the review presentation
      2. Archive procedure
      3. Data Documentation
      4. Data Sharing by NOAA Grantees
   iii. Pending
      1. Data citation
         a. Unique identifier
         b. DOI?
      2. Data Access
e. Resources
   i. This is important work. Need recognition, training. Shouldn't be an afterthough, when-you-have-time job.
f. Architecture
   i. Service-based, system of systems
      1. Community catalogs
      2. Portals
   ii. Archival data centers
   iii.
   iv. Cloud services
      1. Scalability, security
      2. Offload access? Push out disposable copy?
g. Assessment

i. Current state
ii. Progress metrics
    1. EDMC reporting
    2. DM Dashboard
iii. Feedback – users, implementers

h. DM Dashboard
i. Metadata sources across NOAA
ii. Summary reports
iii. Feedback to providers
iv. Publicize good examples
v. Notional:
    1. Accessibility
    2. Documentation

i. Data Lifecycle
i. Planning & Production activities
ii. DM activities
iii. Usage activities
iv. ** directives apply throughout lifecycle

j. Flow – (excellent illustration)
i. Producer
    1. Reqs
    2. DM plan
    3. Gen data
    4. Gen metadata
        a. DOI
    5. Access
    6. Archive
    7. Catalog svc
        a. Dashboard – interface for leadership
ii. User
    1. Create results
iii. NOAA Leadership
    1. Dashboard provides visibility, insight

k. Q&A
i. Phil – Unique ID / DOI?
    1. Early FY13?
    2. Can begin assigning before directive is written/official
    3. NASA has an all-you-can-eat license; we'll do that too
    4. Unlimited – should resolve to landing page forever
ii. Glenn – catalog services
    1. Each has ESRI GeoPortal server
    2. GP working group
iii. John Keck – Data access issues: agility vs security constraints
    1. Working to streamline, leverage cloud:

a. Tony Lavoie, Joe Klimavicz
b. Fed effort
2. Software
a. Pre-blessed software for download
3. Shared hosting project starting

iv. Lynda – correlation with NSF DataOne?
1. Not on a formal level
2. Some interactions w/ other agencies
3. EarthCube
4. NEO strategy
5. Closer to a vision than an implementation

v. Ken McDonald - Is documentation/metadata sole province of data producer?
1. Can happen at all levels
2. User annotation?

6. 09:53 – Ted – Documenting NOAA Data
a. Proc Directive
i. On GEO-IDE wiki
ii. What is metadata – different people view differently
1. NCDC = station history
2. Other places, other ideas
b. "Climategate"
i. Exhonerated, but "failed to be open enough about their work"
c. Users differ – orientation, timeframe
i. WX users
ii. Students
iii. Future (climate)
iv. Regan Moore quote…
d. Foundation across NOAA
i. Develop & implement common metadata management tools
ii. Rubrics to estab baseline & monitor
iii. Promote & highlight good examples
iv. Support training
v. Initiate teams to work on "special documentation problems" across offices
vi. Encourage & support participation
e. Line Office / program process
i. Roles
1. Collectors/Providers
2. Standards
3. Stewards
4. Users
ii. Activities
1. Identify expertise
2. Assess

3. Create/improve
4. Publish/preserve
    a. Ongoing input

f. Assess metadata
    i. Age
    ii. Completeness
    iii. Don't just revise, improve
    iv. Metrics
        1. % with scores above 25
        2. Start measuring while bar is low

g. Standards evolution
    i. Basic content | extended content

h. Doc repository
    i. Multiple dialects (stds)
    ii. ISO is broadest std ~= unabridged dictionary

i. Improving
    i. Spiral development

j. Rubrics, metrics

k. Leadership Model – Positive Deviance
    i. Find those moving right way, ignore the others
    ii. Facilitate communication, provide tools
    iii. %%Barbara Waugh

l. Q&A
    i. Granules are forever
    ii. Say "essential," never "minimum," for metadata
    iii. Ensure long term archive/repository for metadata
    iv. DOI – California digital library? Virtually no cost now?
        1. If don't like DOI, use a UID
        2. Most use UID at granule level
        3. DOI can be referenced to a permanent landing page

7. Break
8. 10:44 back in
9. 10:44 – Nancy Ritchey
    a. Pillars of excellence
        i. Pillars a little like stovepipes – tend to stay in column
    b. Activities & Accomplishments
    c. Q&A
        i. Don Collins - Interagency/international collaboration examples
            1. Looking to further such cooperation here
            2. EUMETSAT
10. 10:52 – Don Collins – NODC Archive Mgmt System
    a. Many holdings are one-off, subjected to a less rigorous appraisal process
    b. Format-agnostic data stewardship
        i. Can accommodate any digital file

ii. OAIS model
    1. SIP, AIP, DIP

c. Accession Tracking database – ATDB –
    i. Postgress
    ii. Version controlled / tracked
    iii. Content (note the relational nature… I'm betting these are analogous to entities, and may each be referenced as a FK as needed)
        1. Accessions
        2. People
        3. Projects
        4. Institutions
        5. Countries
        6. Platforms
        7. Seanames
        8. Datatypes
        9. Instruments
        10. Observations
    iv. WAF

d. AIP Structure (canonical form)
    i. NODC_accession_ID/
        1. about/ = NODC documentation
        2. data/0-data/ = original files
        3. data/1-data/ = NODC translations
    ii. all are version controlled
    iii. previous and current versions are maintained

e. DIP – discovery & acces
    i. Meta provides discovery
    ii. Access services fia ftp, http, dap, las, wms,wcs, ArcGIS

f. Ongoing CLASS coordination
    i. Create, publish ISO metadata
    ii. Access now directly to class via FTP, HTTP
    iii. Want other access options in future

g. Volumes
    i. 78 TB locally
    ii. 112,000 individual accessions (AIPs)

h. Automated Ingest
    i. 17,200 AIPs / year
    ii. 15+ operational automated ingest processes
    iii. Others in prep

i. Non-automated ingest
    i. Avg ~ 240 AIPs / yr
    ii. Non-repeating, single time submissions
    iii. Highly variable
    iv. NODC electronic data doc form – elements map to FGDC

v. Online submission system in prep for Fall 2012

j. File-level metadata & data stds
   i. netCDF for data where possible –
      1. enables many other services
      2. Developed templates to encode novel (non-traditional) data types into netCDF

k. Links
   i. www.nodc.noaa.gov//General /NODC-Submit
   ii. Data.nodc.noaa.gov/geoportal

l. Q&A
   i. Jason Cooper – how address legacy metadata
   ii. Phil – value-added info considered part of same accession?
      1. Consider part of same package
      2. Always keep what they were given
      3. Access
         a. entire AIP available
         b. user can (usually) select what download
         c. FTP, HTTP
      4. Also some aggregation/compositing
         a. Ocean data integrated from multiple AIPs, apply same QC, provide a consistent format
      5. Versioning vs accession IDs
         a. When new rec created, it's 1st point / create structure
         b. If change ATDB rec
            i. Version control each time ATDB record saved
         c. If change in file mgmt. system
            i. New version of access package created
            ii. Moves from read-only into transitory ingest storage, edit, new version is then ingested
            iii. Both old & new are available, both are under the same accession ID
            iv. Eric Ogata @ NGDC
            v. Manages only changes from one version to another
               1. Only the new files are stored
               2. Each version establishes virtual links to other, unmodified files
      6. Ted - FGDC supplemental information
         a. If exists, can refactor into new ISO structures

11. 11:30 fire drill
12. 11:45 – Jaci Mize – tools for creating & editing ISO metadata
    a. MERMAid
       i. Good interim step for FGDC CSDGM to ISO
       ii. Transform capabilities
       iii. No native ISO editing

iv. Will not be supported after next year, but safe to use it now for translation & conversion

b. CatMDEdit
   i. Baby-ISO
   ii. Color coded
   iii. Hierarchical tree
   iv. Can autogenerate metadata for some file formats, including images
      1. Shpe file Jpg, others
   v. Catmdedit.sourceforge.net
   vi. Cons
      1. No 19115-2, Biological, NAP
      2. Search or common catalog
      3. No transforms
      4. No XML attributes

c. GeoNetwork
   i. Robust, flexible
   ii. Pros
      1. Multiple views for maintenance
         a. By package (section)
         b. Entire rec at once
         c. XML view
      2. Supports 19115-2, 19110, 19119
      3. Platform independent
      4. Good validation
      5. Geonetwork-opensource.org
      6. Workflows
      7. Batch ingest/export
      8. Search w/ map
      9. Common repository
      10. Multi-user with a server
   iii. Cons
      1. Complex startup
      2. Validation errors tough to suss
      3. No bio, NAP support
      4. Transform engine doesn't support XPath 2.0
         a. Rich F @ NGDC – can work, but needed tweaking

d. ISOMorph
   i. Divided by section,
   ii. Input via plain-english questions
   iii. Xlink support
   iv. Workbooks
   v. Morph converts to XML
   vi. Pros
      1. Free

2. OK validation

vii. Cons
1. No external validation after generate XML
2. Must save externally

viii. Access
1. asd.radiancetech.com/isomorph

e. Geoportal
i. ESRI
ii. Pro
1. Built-in ISO editor
2. Free
3. Some dropdowns
4. Some autopopulation
5. Workflow

iii. Con
1. Does not support full stds
2. Xform cannot support xpath 2.0
3. Cannot edit auto-populated fields w/o editing code
4. Few XML attribs
5. **On import, removes constructs it doesn't understand

iv. Sourceforge.net/projects/geoportal

f. Altova Authentic
i. Pro
1. Skins
2. Free
3. Customizable
4. Any std
5. Desktop or web

ii. Con
1. Skin development externally
2. Doesn't save recs

iii. Ww.altova.com/download_components.html

g. XML editors next
i. This is NOT like using notepad and raw XML
ii. Lots of help in the tools
iii. Well-received by users
iv. Some learning curve

h. XMLSpy
i. Everything you need
ii. Usable by "average" user
iii. + lots of cool stuff
iv. + Version control support
v. + Batch functions
vi. – not free

vii. www.altova.com/xlmspy.html

i. Oxygen
   i. Ted's group, most of NGDC
   ii. +Good nav frame
   iii. +uses CSS (can implement skins)
   iv. – must set up validation/transform scenarios
   v. – not free
   vi. www.oxygenxml.com
j. FGDC metadata editor review on their site
k. Create ISO via transforms
   i. XSLT / SXL
   ii. Crosswalks defines how, transform uses the crosswalk
   iii. Csdgm2iso19115.xslt
l. Docucomp
   i. Component registry
   ii. Record services
      1. Some transforms
      2. Rubrics
m. EPSG Registry
   i. www.epsg-registry.org
   ii. OGP
   iii. Point to existing resources
      1. Overlap w/ Docucomp?
n. Q&A
   i. Don - NCDDC developing an architecture to permit users to select tool of choice, fit them into a broader system
   ii. Ken - XSLT status, improvements
      1. Problem? E-mail Jaci or Ted & they'll fix it
      2. NESDIS IRMT group – 3rd Monday
      3. What about ISO to FGDC ?
         a. Some available for both discovery and full mapping (with some limits)
         b. Geoportal & data.gov now offer ISO support, so less necessary than before
   iii. Lynda - Transforms for 19115-1?
      1. Transforms now in-process
         a. 19115, 19115-2 to 19115-1
         b. Goal – Oct 1
      2. Lots of namespace changes coming
   iv. Phil – XSLT for ISO to DIF?
      1. One from DIF to ISO
      2. None for ISO to DIF – great opportunity
   v. Misc WRT transforms
      1. Phil - %%Suggests common library for XSLT

2. Kathleen – ECHO to ISO xlst
    a. Ben white's wiki?
    b. Barry creating schemas?
3. Dave Cannel (?) from Antarctica is doing work on transforms
4. JPL
5. GeoPortal

13. 12:20 - Lunch
14. 13:05 – Tammy Beaty – ORNL – Mercury (then Viv Hutchinson)
    a. ORNL DAAC – for archive
    b. Multiple Distrib mechanisms
        i. FTP
        ii. Mercury (FGDC-compliant search catalog)
    c. Diverse Goals
    d. What is
        i. Open source
        ii. Standards based
        iii. Service based
    e. Search and Data Access System
        i. Provider creates metadata
            1. Mercury
            2. Or local tools, then harvested?
        ii. Build index of metadata
    f. Two models
        i. Virtual internet
        ii. Virtual aggregate
    g. Custom interfaces
        i. LP DAAC
        ii. Aggregation
        iii. Multiple sources
        iv. Collection focus rather than granule
    h. Probs w/ collecting metadata after the fact
        i. Hard to remember all details afterwards
        ii. consistency in personnel (new grad student vs one who did actual research)
    i. Demo
        i. Search metadata or Search website
        ii. Tools
            1. Search interface (advanced)
            2. Specify data sources
            3. Keywords
                a. Full text, or by field
            4. Date range
            5. Spatial coords / map
        iii. Metadata shown

1. Station (site) list in one listing
    a. ?? How is the site info represented in metadata?

j. QA
    i. Jay Morris – how know when metadata is updated?
        1. Nightly harvest
        2. Dictionary – GCMD keywords
        3. Indexer – SOLR
    ii. Curt Tilmes – further WRT Mercury
        1. Each dataset has DOI
            a. Resolves to their page
        2. Supports DataOne
        3. https://cn.dataone.org/onemercury/
    iii. Misc
        1. SOLR database is Lucene
        2. SOLR used @ JPL for granule metadata

15. 13:35 Viv Hutchinson – USGS core science metadata clearinghouse (mercury)
    a. USGS data lifecycle (proposed)
    b. USGS CSAS Data documentation workflow
        i. Make it easy for scientists
        ii. Tools, training
        iii. QC (live person)
        iv. Clearinghouse (repository, dashboard)
        v. Share/Push/publish
    c. > 12 year effort
    d. Mercury Consortium through ORNL
    e. Open access online metadata repository
    f. 103k+ metadata recs
    g. Datasets, projects, s/w tools, publications
    h. Standards
        i. FGDC CSDGM
        ii. EML
        iii. ISO 19115/19139
        iv. Dublin Core
        v. DIF
        vi. Others can be added
    i. Offers conversion, also retains original metadata format
    j. Users
        i. All over
        ii. Research
        iii. Scientist
        iv. Land mgr
        v. Students
    k. 93 clearinghouse partners
    l. Participate

i. Push XML to WAF
ii. Notify
iii. Set up harvest
iv. Weekly harvest.
1. Old records deleted
2. Full metadata reharvested

m. Addressing QC
i. USGS FORT – parser/validater
ii. Assistance in improving
n. Background thesaurus to help refine, translate searches
o. Can save search as RSS feed, so always up-to-date
p. View full metadata, or acquire data (dependent on provider)
q. Creating Online Metadata Editor (OME) entry tool w/ ORNL
r. Dashboard
i. Leftover from MBII
ii. Now rebuilding
s. Visualizations
t. Relationships
i. Serves records to data.gov
ii. Domain-specific feed to partners
iii. dataONE
1. NSF initiative
2. Global infrastructure
3. Distributed framework
4. Member node
5. Coordinating nodes
6. Investigator Toolkit
u. DataOne – enabling tools
i. Workflow can be implemented
ii. Morpho
iii. Metacat
iv. Others
v. Education & training
vi. Global communities of engagement/practice
1. Libraries, science educators
vii. vhutchinson@usgs.gov
v. QA
i. Jeff DLB – full replication or distributed search?
1. Coordinating nodes harvest, not distributed search
2. Ted points out that SOLR can do distrib search…
a. Some nodes actually distributed?
b. Would that be determined at the node level?

16. 1359 – Ken Roberts – ATRAC
a. Advanced Tracking & Resource tool for Archive Collections

b. Flow
    i. Forms to populate
    ii. Custom to project
    iii. Shareable components
    iv. Generates documents, NOT a DMS
        1. Need external repository
        2. Look at past versions
c. Display
    i. Search
    ii. Grid
    iii. Timeline
d. Export validate input as ISO
e. Not be-all, end-all
f. Uses info gathered during archive process to kick-start a metadata record
g. Vocabs
    i. GCMD theme
    ii. GCMD data centers
    iii. WMO ECVs
    iv. Other GCMD keywords
    v. ISO topic category
    vi. ISO codelist
h. Modes
    i. Create there
    ii. Import (future)
i. Export
j. ATRAC DB
    i. ?? Not a repository, but still maintains record versions?
k. Demo v2.4
    i. www.ncdc.noaa.gov/atrac
    ii. Search
    iii. Timeline
    iv. Sort by any (?) display field
    v. Anyone can request an account (including outside users)
    vi. Project registration
    vii. Project input
    viii. Can import details from another project
    ix. Metadata/Identification/Distribution/Quality
    x. ?? save/submit, but no cancel/undo/revert?
    xi. Repeating fields
    xii. Nice integration of keyword lookups / checkbox
        1. Some appear to require list selection, others can do partial entry/ matching
    xiii. Distribution (repeating entries)
    xiv. Addl sections (repeating)

1. Process steps
2. Reference docs

xv. Link to (easy manual) validation with DocuComp

l. Validation/Rubric
   i. Lineage?
      1. We use LE, not LI
      2. Not picked up?
   ii. Language
   iii. Online resource
      1. Need names, descriptions, functions for all
m. **Lynda: Value add – this exchange is in fact another quality improving spiral
n. **Jaci – need NOAA transform repository
   i. GEO-IDE wiki?
   ii. ESIP documentation cluster
      1. Rosetta stone? Xpaths WRT similar concepts in all standards
o. Katie – timeline, access control
   i. Specify at project level on creation
p. Ted – how many projects in ATRAC?
   i. ~100 now
   ii. Nightly ISO dump into WAF, then run metrics on it
      1. Automatic validation, link checking
   iii. NB: ATRAC xml output is not necessarily official NCDC metadata record

17. 14:30 – break
18. 15:00 back in
19. 15:05 – Lynda Wayne – FGDC Metadata Summit
    a. October 2011 @ USGS HQ – Reston
    b. USGS, NOAA, Lynda (FGDC)
    c. Training, Summit
    d. 52 participants, 25 agencies
    e. 3 groups
       i. Implementation
       ii. Tools
       iii. Communications
       iv. <standards – no one signed up>
    f. Recommendations
       i. Policy/Guidance
          1. Highlight/demonstrate benefits
          2. Implementation guidance, use cases, workflow solutions
          3. Draft/promote directives specific to ISO imp
             a. Some reticence unless forced to
          4. Tie implementation to agency mission, activities, etc
       ii. Education & Communication
          1. Multi-level, role based training

2. Offer via webinars & recorded sessions
3. Continue classroom training with hands-on
4. Establish online help desk & workspace
5. Provide clear info about stds process & status

iii. Tools & Apps

1. "implement the technology so that people can do the work"
2. Update FGDC editor review
3. Work with vendors to fully implement
4. Work with ISO fo finalize
5. Develop portals to publish ISO meta

g. FGDC Support Strategy

i. Policy

1. How to grow teeth given existing DNA

ii. Guidance

1. Downloadable pubs synced with web content

iii. Training

1. Train the agency metadata expert

iv. Tools

1. ISO explorer? Replacement to greenbook workbook

v. !!! Communications via MWG

h. Workplan

i. lwayne@fgdc.gov

j. QA

i. Ted - Retrain on translation / conversion

ii. Leanne Cross, Woods Hole – verifying result of a translation is what was intended

1. Concern with blindly follow translation, but no basis for review if one doesn't understand the target standard
2. Ted will introduce tool tomorrow (Docucomp?)
3. Jaci - Online recordings for training are available

iii. Ted suggests EDM wiki as best location for best-practice use

iv. Jaci – CAP grant should again be available to fed agencies, since much of the work is done there

20. 1530 – Scott Ritz – GCMD science team lead

a. Latest GCMD released June 12

i. Largest release in memory

ii. Major backend changes

1. KMS – keyword web service
2. MWS – metadata web service, RESTful queries
3. NextGen – preview of new web site (delivered this fall)

iii. Demo

iv. Usage Metrics

v. Feedback

b. KMS

i. RESTful
    1. SKOS – simple knowledge organization system
    2. RDF – resource description framework
    3. OWL – Web Ontology Language
    4. CSV
ii. REST paths
    1. Concept resource
    2. << others from presentation>>
iii. Now UID, so always a stable identifier

c. MWS
    i. List of all docs, filtered by query exp
    ii. Specific doc, via UUID
    iii. Unique IDs of metadata docs
    iv. Keywords for specific field
d. Authentication for services using EOSDIS user registration system
    i. urs.eosdis.nasa.gov
e. NextGen site
    i. New design/structure
    ii. Improved access
    iii. Integ w/ MWS/KMS, docBuilder
f. Demo
    i. Gcmd.gsfc.nasa.gov/nexgtgen
        1. Datasets
        2. Services/tools
        3. Ancillary descriptions
    ii. Can refine searches
    iii. Filter, or can traverse keywords
g. About – complete description of new features
h. QA
    i. Katie – Tech platform?
        1. Java
        2. Spring
        3. Google Web GWT
    ii. Wendy – Latest DIF spec?
        1. Site map . writer's guide . XML DIF Schema
    iii. Phil – DIF changes?
        1. Latest was to add dataset DOI
        2. Infrequent, since it can break things
    iv. Phil - WAF harvest planned?
        1. Use DocBuilder, or
        2. Bulk upload (contact, get Scott to pull)
        3. CLASS
            a. Scott pulls FGDC (No ISO to DIF transform)
    v. Ted - Capabilities document for services

1. From the "about us" links on new release

21. 16:02 – Ken McDonald – CEOS CWIC
    a. CEOSS WGISS integrated catalog
    b. Co-leads
        i. Yonsook Enloe
        ii. Martin Yapur (on phone?)
    c. GEO
        i. Began 10 yrs ago, GEOSS 10 year plan
        ii. Ministerial level, multiple countries
        iii. Secretariat in Geneva
        iv. http://Earthobservations.org
    d. GEOSS Common Infrastructure defined
        i. Results from registry search often too broad, uneven to be fully useful
    e. CEOS
        i. 1984 - Predates GEO 20 years
        ii. Focus on satellite data
    f. WGISS – working group on info systems & services
        i. Http://wgiss.ceos.org
    g. Directory/Inventory search
        i. IDN – int'l version of GCMD
        ii. Each CWIC collection – unique DIF
        iii. CSW (catalog service for the web) inventory, search protocol
        iv. Need both data center name & CWIC collection to initiate inventory search
    h. Uses wrappers/connectors to ensure query works with partner/provider API
        i. CLASS – early adopters of the NEAT interface (no longer supported for new datasets) New API in development
    i. http://wgiss.ceos.org/cwic
    j. International participation is a particular strength, fosters adoption/ implementation of standards
        i. Esp OGC CSW 2.0
    k. Still proto, pre-operational soon?
    l. QA
22. 1631 – Curt Tilmes – US GCRP / NCA / GCIS
    a. 1990 – GCRA passed
    b. Global change ~= climate change + impacts
    c. Tom Karl chairs subcommittee on GCR
    d. http://assessment.globalchange.gov
    e. Major issues
        i. Transparency, authenticity, traceability
        ii. Utility, integrity, objectivity
    f. "traceable accounts" - Contemporaneous notes, summarized, would go a long ways here
    g. http://nca2009.globalchange.gov

h. Authoritative, accessible, usable, timely
i. Stable, unique Identifiers, with traceable provenance
j. QA
    i. Good luck with that
    ii. Ted – DOIs
        1. USGCRP could buy DOI license & make available to all 13 participating agencies
    iii. Timeline
        1. Report released Dec 2013
        2. Datasets need DOIs as soon as possible

23. 1659 – Christina – wrapup

Wed 8/15/2012

1. Attendees
    a. Jeff De La Beaujardière
    b. Yonsook Enloe - NASA
    c. Ken McDonald
    d. Glenn Rutledge
    e. Ted Habermann
    f. Nancy Ritchey
    g. Jaci Mize
    h. Don Collins
    i. Luther Lighty – NASA/Echo
    j. Kathleen (Katie) Baynes – NASA/Echo
    k. Phil Jones
    l. Valerie Toner
    m. Heather Brown
    n. Ken Roberts
    o. Curt Tilmes – GCIS
    p. Ana Pineiro Privette
    q. Linda Copley
    r. Lola Olsen
    s. Scott Ritz
    t. Jay Morris
    u. Tammy Beatty…. – ORNL
    v. Christina
    w. Lynda Wayne
    x. <John Keck>
    y. <<others from NCDC in back rows>>
    z. <<phone-ins>>
2. 08:15 – Jay Morris – CLASS (in-progress)
    a. Access <> Dissemination
    b. CLASS concern not to offend/compete with data centers
        i. Requires little metadata for its activities
        ii. Not intended to be metadata catalog, or provide rich search functionality
        iii. Search/Discovery, particularly for in situ / data center data, is province of data centers
        iv. Satellite access built on reqs of satellite data providers
            1. Machine to machine interface
        v. Granule vs collection metadata
        vi. Need to ensure metadata is synchronized, esp when mods/ corrections are made
        vii. Minimize number of archives across NOAA, consolidating where possible
        viii. components
            1. Common storage service – “cloud access pilot”

a. Often used / new data pushed to spinning disk for quick access
b. Push once, read many
2. Common ingest
3. Common archive
4. Satellite ingest/storage/archive
5. Subscriptions

c. QA

i. Yonsook – granule metadata specified from science perspective
1.

ii. Mike P – how are versions handled?
1. CLASS does not have concept of version
2. New version = new object
3. Particular need for metadata versioning
4. How implement versioning w/o multiple copies of same object, but maintaining audit trail?

iii. Ted – follow on WRT versioning
1. Everything really needs an ID to permit linkages/relationships to be documented.
a. Data family
b. Group
c. Dataset
d. Instrument / platform
e. Version
f. Etc
2. Which side of "magic yellow line?"
a. Stewardship / data center issue
b. Archive is just a store and retrieve
c. Suggest: when a new version created, a new ID assigned. Job of data center to track it.
3. How archive / backup catalog? Add to archive? Recovery, long-term survivability independent of technology/platform (i.e. Oracle goes away)
4. Danny – must maintain relationships to permit reassembly off archive
a. Example:
i. Archive packet has 50 stn obs
ii. 2 stns change
iii. New version: archive whole package, or just the new stuff and then reassemble based on noted relationships
5. Jeff DLB – discussion @ dinner re: moving from big black box to a modular, service-oriented architecture

a. Need better definition of how functionality is partitioned between CLASS, data centers, elsewhere
b. Philosophical discussion, can be very political
    i. Who will be catalog of record?
    ii. One catalog or many?
    iii. Who provides the cloud functionality?
c. Can do:
    i. Describe info needed to put info into archive & retrieve it
    ii. Define & expose interfaces

6. Mike P – constant new in situ data… what is reasonable amount of data to store in a single file-level granule.
    a. Depends on data center
        i. Some want to provide a few large TARs
        ii. Some want to provide all obs individually
        iii. Alan – want to avoid data package so small that the metadata is larger than the actual data
        iv. Ted – disagree WRT size of metadata vs data
        v. Clarification WRT "metadata" – granular, file-description metadata

iv. %%JDA – sounds like in essence CLASS is a hard disk with FAT table. All structure & logic is virtual, imposed by provider

3. 08:50 – Katie Baynes – ECHO
    a. Based at Goddard
    b. EOS – Earth Ob System collect Earth remote sense data for glocal change research program
    c. EOSDIS (EOS Data and Info System)
    d. Distrib centers of expertise
    e. ECHO is the service-based middleware layer between data & users
        i. Dev as part of NASA's core data sys
        ii. Model since 2007
        iii. Fully adaptable
    f. REVERB - NextGen web client
        i. http://reverb.echo.nasa.gov/reverb
        ii. Query UI
        iii. Tools to examine results
        iv. Expose service registry (M2M via REST)
        v. ECS-style data partners
    g. Definitions
        i. Providers – mostly DAACs
        ii. Collection – group of related granules or ISO datasets; common attribs about granules
        iii. Granules – indiv spatial or temporal data elements (ISO datasets)

iv. Resources - Referenced by granules or collections; browse metadata, for example
v. (Provider 1:M (collection (1:M granules 1:M resources) 1:M resources))

h. Provider metadata ingest to ECHO catalog
   i. Formats: Echo 10, ISO 19115
   ii. FTP ingest, or RESTful API
i. Discussion: interactions between catalog, archive
   i. Tammy – can be frustrating if GCMD sends user to ECHO which then sends user to CLASS (or elsewhere)
   ii. ISO 19115 providers
      1. CDAC
      2. Internal NASA project
j. Can do stratified/restricted access by user group
k. Docs at http://api.echo.nasa.gov/catalog-rest/catalog-docs/index.html
l. ESIP connector (OpenSearch)
m. Metrics
   i. 8-12k queries / wk
   ii. Performance has sometimes been an issue, but usually < 10 seconds response
   iii. Goal 99.9% uptime
   iv. 2TB index covering PBs of data
   v. Public holdings
      1. 2798 collections
      2. 105 million granules
      3. About 106 million granules indexed,
      4. grow by ~100-200k / wk
   vi. Other
      1. 60k+ registered users
      2. 12 active partners
      3. 34 mil ECHO-hosted browse images
      4. 11 operational clients (??)
      5. Several clients in T&E
n. Three environments
   i. Operational (beefy, balanced)
   ii. Partner-test (more modest, some virtualized)
   iii. Testbed (REST only)
o. QA
   i. Working on LANCE(?)/MODIS ingest
   ii. Phil – of 100k, proportion of replaces to new
   iii. John K – how handle scalabilty
      1. Working on it
   iv. John K - Team size (total)
      1. ~15 people, and some shared with other projects
         a. Agile, with 2-3 wk sprints

b. 3-4 devs
c. 5 essay team
d. 1 lead
e. Mgrs.

2. Timeline?
   a. 1st Proto x years ago
   b. ECHO 10 – 2007

p. Back to presentation
   i. Embrace change
   ii. Limitations
      1. 19115 is a snapshot
      2. Future planning
      3. Validation
      4. Service integration

q. QA
   i. Jeff DLB – data formats mean metadata formats
      1. ?? order process
      2. Put order in cart, assign UUID
      3. Hand-off to provider
      4. Data center provides status update to ECHO, provides user w/ link to result
      5. PUMP – permits authorized users to make some updates?
   ii. Ge Peng - Subsetting
      1. Interfaces to subsetters, but don't do subsetting
      2. EOSDIS service interface
      3. ??visualizations?
         a. Overlays of extents (granule level) on REVERB map
   iii. Ken – limited support for ISO
      1. Pull XPaths, based on 2009 spec
      2. Cherry-pick "core" ECHO fields
      3. Do this for collection & granules
      4. Ted – this is same "extract & index" concept that GeoPortal, Mercury others use
   iv. John K – use cases / users?
      1. No direct contact
      2. Volunteer user survey
      3. They work w/ DAACs to develop user profile, problems
   v. Steve A - Interact w/ Giovani
      1. Giovani is client, uses OpenSearch interface

4. 0940 – Ted – Docucomp
   a. Dialects & mapping
      i. ECHO has some info beyond 19115
      ii. Pending 19115-2 revision may (will?) include some of these features
      iii. %%Ted is good POC for input to revision

1. Last round, 700+ comments to review w/in 8 hrs
2. NOAA participation important

b. The Road to DocuComp – overview of drivers
c. Similarities to Normalized Database model
    i. Xlinks (like FK) to reference components
d. OpenDAP
    i. Get data out of files, into clients
    ii. Heavy lifting is in xlating from file systems & databases into DAP to pass to client
    iii. w/ metadata, XML fulfills the DAP role
e. Resulting XML is view of content based on resolved keys
f. Roles, Identifiers, References
g. Separate reference from object
h. Jane as a reusable component
i. DocuComp = Database + RESTful service
    i. http://ngdc.noaa.gov/dolcucomp/<<UUID>>
    ii. Returns XML representation of component
j. Resolution Flow
    i. RecordSet/iso_u/xml = unresolved
        1. _u = unresolved
    ii. RecordSet/iso/xml = resolved
    iii. Original plan – just expose resolved
    iv. Evolving req – distrib w/ references unresolved
        1. Compact
        2. Citations particularly bulky
k. Getting there
    i. FGDC WAF =(harvest)=> FGDC @ NGDC =(translate)=> ISO_U =(resolve)=> ISO
    ii. Translation proofing
        1. Some limitations on translation
            a. Address type ain't happening
    iii. Rubrics
    iv. Validation
    v. Link checking for online resources
        1. Xlink checking
    vi. Multiple views
    vii. Consistency checker – find repeated content & candidate components
    viii. Docucomp to create/store components
l. NOAA EDM wiki documents all this, ISO explorer and more
m. Good example of integration, linking: Ken demo ATRAC linking to wiki for code lists
n. Consistency Checker
    i. Visual inspection

ii. Reports based on automated evaluation rules

o. QA – Discovery FGDC? links back to original ISO rec?

p. QA – Jeff DLB - how create components not based on ISO std? ex: SensorML description of instrument

q. <<Potential versioning issues for unresolved records, since component could be updated unbeknownst to metadata record owner>>

r. Lots of resources on wiki

s. www.ngdc.noaa.gov/docucomp

t. QA / comment

i. Jaci – unresolved records are dynamic. Must archive snapshot of resolved record for versioning

ii. Steve - Revision history for components?

1. No, but resolved records are in SVN

iii. Ge Peng – validating components

1. Can put blobs of anything into the component
2. Validated on resolution

iv. Jeff DLB – provide validation schema for non-ISO components?

v. Phil – permissions on component editing

1. Yes, now
2. Users/owners to enforce privs
3. Used to be able to directly access via Oxygen, XMLSpy, but security concerns limit that
4. Need to validate through oxygen client

vi. Jim B – NGDC will have DocuComp in perpetuity?

1. UUID is portable
2. Could stand up a new service, redirect, but would take some manual handshaking to make the transition

vii. Steve A – embed components in NetCDF files, as do for ISO

1. Direction they are heading
2. Xlinks not supported by ncML, so some limitations in std
3.

5. 10:17 – break
6. Jaci – ISO implementation, best practices

a. Working from GoogleDocs "Discussion Points"

b. Overview of tools & resources

i. Demo is live

ii. %%JDA – get bookmarks from Jaci

iii. **Centralized location for transforms?

1. EDM wiki?
2. GitHub?
3. ESIP wiki?
4. SourceForge?

iv. NCDDC website – ftp.ncddc.noaa.gov/metadata

v. Altova

vi. ncISO
1. part of thredds
2. converges CF, THREDDS with ISO
3. working with unidata for providers to provide own XSLs

c. Automation
i. Creation & maintenance
1. both can benefit from automation
2. Needs may differ

d. ** discover, analyze what's available before rolling your own
e. ** communication is key
f. Implementation
i. Full life cycle approach
ii. Standardize usage, vocabs, validation, handling versioning, etc

g. Challenges
h. Don C - NODC
i. Controlled vocabulary cleanup
ii. Mapping content to XML
iii. Developing workflow to accommodate additional steps
iv. Training

i. Edit=>translate vs translate=>edit
i. Case-by-case basis
ii. Differentiate between edits needed for translation, vs edits to improve metadata

j. **Ted - Translating existing records to ISO is a level 0… needn't improve to get it translated as a baseline
k. Political/sociological changes are more difficult and complex than technology changes
l. Lynda: How to move forward? What are first steps
m. Lola: controlled vocabularies are critical
n. Ted: terminology – "shared" vs "controlled"
i. Easier sell

o. **Again, ensure optional "nice to haves" aren't presented as requirements for migration to ISO
i. Migration vs improvement

p. Scott – preserving integrity of record during translation between standards
q. Ted – must quantify translate shortcomings
r. CSDGM – ISO
i. Address type
ii. Dial-up phone number
iii. Baud rate?? (joke)

s. Must eschew different flavors to ensure interoperability
i. NASA-flavor of ISO?

t. Lynda – extension vs profile
i. Extension adds elements

   ii. Profile specifies how to use existing
u. Jaci – schematrons
   i. Check validation against schema
   ii. Schematron is broader, can check content, not just syntax
   iii. Use those to enforce usage profiles
v. Customization – skin can restrict which domain for picklists
w. ** also centralized shared skins, not just XSLTs
x. Tammy – see editor as series of controls that string things together for consistency
   i. Checkout w/ version control
   ii. %%I think this is more workflow than editor. Process vs. tool
y. On EDM wiki – NESDIS Enterprise Documentation System
   i. EnterpriseMetadataArchitecture.png
   ii. Illustrates workspace vs published subversion repository
z. Phil – defining workflow to ensure records are reviewed before published
   i. NCDDC –dashboard controls workflow for approvals
   ii. Workflow implementation (Erik Robe question)
aa. Ted – NGDC, each group has own review/approval process
   i. Concern that approval not become a bottleneck
   ii. Was dissatisfaction at NGDC
bb. Jeff DLB – Mandate ISO metadata record as the required exchange format
   i. Require ISO record be in ISO WAF before it will appear in portal, etc
   ii. Data.gov can now ingest ISO recs
cc. Don – when/how is something published
   i. Metadata – internal database; updated, pushed to WAF whenever metadata is updated
   ii. Separate from making AIP discoverable, accessible
dd. Katie – ECHO handling quality & versioning
   i. Talks w/ LANCE-MODIS
   ii. 3 hrs fresh
   iii. Rough processing before distribution
   iv. Interim acces. Propose to age out @ 2 wks.
   v. Another data center later provides access to more refined, QCd version
   vi. Jaci – can key off of revision date, status (in process, complete, etc)
ee. Lola – all speak English
   i. Implications for international exchange
   ii. Keywords help standardize, since 2nd language speakers can recognize words more readily than think of them unprompted
   iii. Jaci – code lists help address this directly
ff. Yonsook – GOES-R metadata found codelists very limited
   i. After eval, GCMD keyword lists offered richer representation
   ii. Added keywords to GCMD list (with coordination from GCMD, as the controlling authority).

iii. %%JDA – implications for codelist versioning WRT metadata record version

gg. Ted - Some believe ISO should have no codelists at all?
- i. Want to insulate standard from need for change
- ii. No codelists vs some codelists
- iii. Mapping / xrefing codelists
- iv. SOLR Facet search shows multiple synonyms

hh. Katie – ECHO tech committee
- i. Requests to remove ECHO faceted search
- ii. Ex: Processing level confuses users (because providers are inconsistent in what they provide)
- iii. Issue: keywords are whatever providers send, and they want to retain their flexibility / freedom

ii. Ted – data.gov analysis - inconsistencies w/ agency names
- i. NOAA, USGS have many, many synonyms
- ii. Chris's blog post, data.gov
- iii. Use schematron to ensure items are valid codelist entries, where codelist values should exist

jj. Phil - Consensus for GCMD keywords to use both short & long name?
- i. Depends on keyword list
- ii. Suggest using entire string, with shortname & longname

kk. Phil – current use case to justify continued use of FGDC
- i. No. no reason not to migrate to ISO & use it

ll. FGDC Entity & attributes
- i. Now point to externally
- ii. Next revision, can bring it inline
- iii. Some have used MI_band for E&A?

mm. Ted - History of ISO std dev, why things are what they are
- i. Military drove much of input for first rev of std
- ii. NOAA, NASA provided less input, thus their needs not as well met
- iii. Chg name of MI_band to MI_dimension (?)
- iv. Prob w/ feature catalogs vs coverages
  1. GIS – vector vs raster
  2. Coverage and feature are kind of synonymous
  3. ** recommend using coverages instead of features

nn. Issue: versions of standards differ; which to conform to?
- i. Use future/draft/pending vs current version

oo. CF Convention names – standard vs non-standard

pp. Phil – use ncML instead
- i. Added record, record type into content info
- ii. Can use to embed ncML into ISO
- iii. Examples on wiki
- iv. They do same with ECHO product-specific attributes
- v. Dealing with record/record type is harder than fudging ones definitions

qq. Discussion, concerns with stability of standard
rr. Guidance: build to 19115-1 draft standard
    i. 19115
    ii. GML (from OGC)

7. EDM wiki includes -1 discussion
8. Services
    a. If humans read it, put it in distribution
    b. If machines read it, put it in services
9. Citation
    a. Cite up, from child to parent, rather than down, from parent to child
10. Association type
    a. Ted suggests using lineage
    b. Jaci – aggregation used for "who funds"
11. 12+ - Lunch
12. 13:15 Back in
    a. QA & discussion
    b. Phil – Item L – XSLT to HTML view of LE_Linkage?
        i. Transform to GraphML, which outputs SVG, which can be displayed in browser
        ii. Done for all NESDIS datasets in SPEEDS
    c. DOI discussions
    d. Curt –
        i. DataOne wants DOI to resolve to a landing page, not necessarily the page that leads you directly to the data
        ii. Versioning – should each version of a product get its own DOI?
            1. Curtis, Jeff DLB both say separate DOI for each version
            2. Should there be a common landing page, one for each version, or ??
    e. John K – create common Drupal data types to group similar data
    f. Curt – DOI makes a good, linkable reference
        i. How can you make the landing target useful for human readers and also for machine?
        ii. Jeff DLB – make target XML, use XSL to put in appropriate format for human or machine depending upon what is accessing it (overloading)
    g. Ted – Implications for identifiers of subsetting
        i. people usually use only part of a dataset
        ii. OPENDap permits subsetting
        iii. Review logs, use the openDAP link
        iv. DOIs are the flavor-of-the-month, but there are other issues, and approaches, that remain
        v. ARK identifier?
            1. https://wiki.ucop.edu/display/Curation/ARK
            2. Jeff DLB says used by French digital library
        vi. DOI is too general for detailed, reproducible provenance

vii. ** possible IRMT discussion on this topic?

h. CLASS is implementing its own identifiers
   i. Use UUIDs instead!!
   ii. !! How resolve?

i. ** General consensus (?)
   i. DOI at dataset level
   ii. UUID at granule level
   iii. In context of scientific data citation, how should these be used together??

j. Jeff DLB – need a persistent identifier;
   i. still have opportunity to select something besides DOI
   ii. NCA needs to move forward, so we really do need to make a decision

k. Jim B – what is actual utility/need for UUIDs
   i. Ted - Large volume issues, 20k granules
   ii. Lineage / identification

l. Curt – mentions Duerr article on DOI/ARK/UID in context of scientific data citation

m. %%JDA – get link to article from Katie

n. John – need to prioritize the ID issue based on scale, immediate need, complexity

o. GRIST??

13. 13:48 – Scott – Metadata Use
    a. Discovery
       i. Datasets - DIF
       ii. Services/tools – SERF
       iii. Ancillary Descriptions –
    b. Climate diagnostics - visualization
       i. NOAA request
       ii. Via CEOS?
    c. Adding descriptive information
    d. Tool - DocBuilder
       i. EOSDIS account
       ii. Create
       iii. Modify
       iv. DocumentIdentifier
           1. Must be unique
           2. Otherwise, totally freeform(?)
       v. Public production system vs cache, which is offline not publicly available
           1. Cache copy available 90 days, for revision, limited sharing review
       vi. Can edit fields or edit XML directly
    e. Sharing / Registration
14. QA ensues
    a. Lynda – metadata use goes way beyond discovery
       i. Litany of "A" words
           1. Access

2. Assess
3. Archival
4. …

b. Jay – portals (archives, too) can require adequate metadata before access will be provided
   i. Discuss required vs recommended
   ii. Yonsook – addressing some of this in CWIC
   iii. Lynda – FGDC's "essential metadata" document
c. Ted – role of users in metadata
   i. Usage stories
      1. who tried to use the data, how
      2. what limitations encountered
      3. Response from provider
   ii. John Bates – CEOS climate visualization
      1. Tool finds stuff
      2. Limited metadata results in insufficient/incorrect attribution
   iii. Dangers of making poorly documented data discoverable
   iv. John B – example from Climate.gov
d. *** how to encourage/motivate/force data providers to include "enough" detailed metadata to permit not just discovery, but lineage, attribution, assessment, interpretation, etc
e. ESIP rant – what I could do if xyz metadata info were available (Dave from USGS)
f. Ge Peng – additional metadata flags for usage?
   i. <<Not sure I understand the question>>
   ii. Lola responding; suggests send recommendation to her

15. 1615 - %%Broad (and somewhat rambling) discussion ensues
16. 1626 – Scott (continues)
    a. Necessary for sharing
       i. Stable, unique ID
       ii. Versioning
       iii. Ted – record orientation can be limiting
          1. Faceted searches
          2. Idea for metadata editor to help unify free text entries: option to display existing values in a given field across entire collection
          3. Applies to versioning
       iv. Phil – versioning is a data producer issue, often dependent upon internal procedures and affected by their business rules and how the dataset is produced. Thus, probably out of scope for these discussions
    b. Harvesting
       i. Shaida – harvesting techniques need to be well documented & made available to submitters. Harvester may have specific requirements based on how they use the metadata
          1. Geoportal – problem was with content

ii. Alternative – dynamic queries across servers
    1. Federated search
iii. Overwrite/replace vs incremental harvest
iv. CSW –
    1. GCMD uses IDN CSW server (GeoNetwork catalog)
        a. Three servers running: production, failover, testbed
    2. Geoportal (more stable)
    3. Geonetwork
    4. Initial concerns with stability of std implementation
    5. Rich – breaks us out of clearinghouse mode, reduces harvest requirements
v. Yonsook – 2 clients using CSW
    1. Ratheon effort – based on REVERB source code, refine, search CWIC
    2. LSI portal – search CWIC
vi. CSW partners
    1. GEOSS/CI
    2. CEOS/CWIC
    3. ERO-GEOSS
    4. Climate.gov
vii. Tweaking / caching

c. Geoportal configs
    i. XPaths (Jaci)
        1. easier to implement if use // rather than explicit paths
        2. Not a true lucene index – make a body tag as a wrapper
        3. Documented on EDM wiki
    ii. Work w/ NGDC on onlinelinkage
        1. Type – informative, website, bulk download, metadata
    iii. Geoportal evolving. Posting config stuff on wiki to help foster consistency
    iv. Rich – dependencies for Geoportal's use of onlink tag
        1. Limited to two initial
        2. Order added to permit prioritizing
        3. App profile to denote THREDDs , FTP, etc by content type rather than order
        4. Customization needed to take full advantage

d. Granule metadata standards
    i. CF conventions
        1. Standard names
        2. Usage conventions
        3. Enables to open file, make a picture without knowing anything else
    ii. ACDD –
        1. Unidata proposed
        2. Standard names for attributes for discovery

iii. How to map some unique DIF stuff to CF
iv. Application Profile – ISO component that doesn't have a direct analog in other standards
    1. GCMD 1st group that recognized the need
v. Need: support to move groups concept forward in netCDF
vi. Problem: role name is in the standardized name, rather than associating a party with a record by role
vii. Ted's heading ESIP documentation cluster
    1. Will be sending CF recommendations to committee
viii. Shaida – granules – quality flags
    1. Each sensor has separate quality flag
    2. No standardized approach to quality flags
    3. Curt – community was addressing this under Quality cluster of ESIP
        a. Brent Mattox (?) leading
    4. ISO – 19157 standard for data quality
        a. No schema
        b. Primary standard writer retired
        c. Need a schema written

17. 15:00 – Break
18. 15:20 – Ted – Tools
    a. How do organizations work? Why do they sometimes not?
        i. Creative people in noncreative environment
        ii. Lack of tools
        iii. Lack of organizational support
    b. Tools
        i. FGDC
            1. Initially no XML
            2. Initially no tools
        ii. Can't compare early days of ISO to early FGDC, because there are now tools & techniques that may be brought to bear
        iii. Management tool may include
            1. Schema
            2. Schematron
            3. XSL
            4. CSS
    c. Who should be creating metadata?
        i. Not scientists?
        ii. Model – library
            1. Metadata for books not created by author
            2. Difference – does librarian actually document content of book?
            3. **Data curators
        iii. Christina

1. Need tools so that scientists can provide information about their data
2. Scientists provide content in plain English
3. Metadata technician takes content and generates / refines into good metadata record

iv. Jaci – suggests that with proper templates, scientists in fact will (may? Can?) generate their own metadata
v. Use ATRAC to get scientists involved
vi. Lynda – integrate tools, and the metadata creation PROCESS, during and early in product development lifecycle

d. <<Things is getting a might heated here – not sure why>>
e. Discussion of tools vs procedures vs people (facilitation, training, support)
f. Focus on outcomes
   i. What do we really need?
   ii. How do we meet the needs?
   iii. ATRAC helps implement a desired outcome via a defined process
g. <<things still warm, but everyone's calming a bit>>
h. Jaci – focus on functions & outcomes
   i. Modular functionality
   ii. String them together, don't reinvent
i. Ted – implement ATRAC as a skin on top of some other tool?
j. Jim B – XML, etc is a just representation of content.
   i. Content is the ultimate focus
   ii. Important to abstract the concepts and content
   iii. XML, tools support managing the content
   iv. Are metadata techs like secretaries?
k. Jeff DLB – tools are not so much the issue. Need to focus on automated generation of metadata

19. 16:15 – Back into actual presentation
    a. XML management tool (I think the stalking horse example of this tool is Oxygen)
       i. Technologies
          1. Schema
          2. Schematron
          3. Xsl
          4. Css
       ii. Standards
          1. Iso
          2. Fgdc
          3. Echo
          4. Dif
          5. Thredds
          6. NcML
          7. EML
       iii. Supporting stds

1. OGC services
2. SensorML
3. DocBook
4. UML, Office
5. RDF, OWL

iv. Versioning
1. Subversion

v. Content
1. Underlying model needs to handle all content needed by all standards

b. <<Is this the antithesis of the agile, modular service approach to functionality, or is Oxygen merely the integrator of such services?>>
c. Wiki
d. contact Ted for account on metadata wiki
e. correct wiki and then send email to Ted (or just send email to Ted)
f. Rich: much netcdf doesn't have complete set of attributes, how do we do quality control?
   i. - nciso has rubric for comparing netcdf to thredds, cf, acdd
   ii. - has been in last version of Thredds (Ansari aware of it)
g. Challenges in auto-generating adequate metadata
h. Ansari: manual vs. auto-gen'd metatdata record
   i. - dream aggregate metadata and auto-update metadata record
   ii. Ted: problem is acdd is discovery convention, bare bones; iso rubric is 0-40, acdd conversion will rarely get above 14
   iii. given high-quality netcdf - you can write high-quality iso, acdd or cf
   iv. ted: if you can get really high-quality metadata into the granules, then its stil there 20 years from now!
i. <<back in – Thanks, Linda!>>
j. 5-10 year outlook?
   i. NetCDF vs HDF?
   ii. netCDF orientation is write-once, read-many
   iii. HDF group more amenable to ongoing update
k. Yonsook – collaborating to better define netCDF NcML metadata
   i. NGDC
   ii. NCDC
   iii. NcML templates

20. 16:55 – Christina – wrapup

Thu 8/16/2012

1. Attendees
    a. Jeff De La Beaujardière
    b. Yonsook Enloe - NASA
    c. Ken McDonald
    d. Ted Habermann
    e. Nancy Ritchey
    f. Jaci Mize
    g. Don Collins
    h. Luther Lighty – NASA/Echo
    i. Katie Baynes – NASA/Echo
    j. Phil Jones
    k. Valerie Toner
    l. Heather Brown
    m. Ken Roberts
    n. Curt Tilmes - GCIS
    o. Ana Pineiro Privette
    p. Linda Copley
    q. Lola Olsen
    r. Scott Ritz
    s. Jay Morris
    t. Christina
    u. Tammy Beaty
    v. John Keck
    w. <<others from NCDC in back rows>>
    x. Rich B
    y. Neal Lott
    z. Axel
    aa. Ge Peng
    bb. Erica Johns
    cc. Jason Cooper
    dd. <<phone-ins>>
2. 09:00 – Christina - discussions
3. Beneficial? Continue? How make things work better?
    a. Lola – good blend of viewpoints, agencies
    b. Don – interesting, useful, but not sure of task goal
    c. Already some venues / groups
        i. IRMT Metadata Monday
        ii. Various center working groups
    d. Jaci – quarterly IRMT briefings for various local groups?
    e. ** Jaci, Kathi Martinolich are both tasked as “NOAA Metadata Support”
    f. Communication
    g. EDMC wiki?
        i. Open read, need account to write

        ii. Ted creates accounts
4. Ted -
    a. ESIP clusters
        i. Interoperability
        ii. Documentation
        iii. Data Stewardship
    b. IRMT
    c. Data Documentation directive
    d. Reusable components
    e. Hierarchical docs/metadata
    f. Granules
    g. Datasets
    h. Services
    i. Resource lineage
    j. Data quality
5. ESIP overview & discussions
    a. Esipfed.org
    b. Recent hiccup in NESDIS funding; looking for broader NOAA involvement
    c. Wiki
    d. Commons (drupal)
        i. DOI assigned to what's posted
    e. More general training, geared to different audiences
        i. High school education
        ii. Data mgmt
            1. Still need modules, reviewers
    f. Clusters are listed on site
    g. Several meetings collocated with DataOne
    h. NSF EarthCube
6. Jeff DLB
    a. Estab. NOAA metadata mailing list
    b. Wiki is good, but limited since
    c. DMIT
    d. NESDIS IRMT
        i. Some non-NOAA people
    e. Tiger Teams as needed
    f. Focus on ensuring that every NOAA dataset in NCA is well documented
7. Yonsook
    a. Granule metadata, esp WRT GOES-R
        i. Rev C model is here now
        ii. Rev D Nov 2012
        iii. Final June 2013
    b. Active input from NOAA, including Phil & Axel
    c. GOES-R stakeholder group Tue/Wed 2-4??
8. Groups/Teams

    a. NCA
    b. CLASS/GOES-R/NASA-ECHO
    c. GCMD
    d. GOSIC
    e. Climate Portal
    f. Keywords
9. Christina
    a. Know what's out there
        i. Work together
        ii. Adopt / adapt, don't reinvent
    b. Communicate
        i. Wiki
        ii. Mailing list
10. Phil – How involve data producers?
    a. Must push requirements upstream
    b. Yonsook – WRT satellites, MUST get reqs in early, well before mission launches
    c. EDMC – Data Management Directive
        i. Darien Davis, OAR – spearheading OAR DMP effort
        ii. Ken Masari (?) – WMO WWW; attribution for data distribution
    d. Incentive vs Enforcement
11. John – need to bound problem
    a. Each term/concept is a full project
    b. Looking for big picture…
    c. Focus on most important aspects
    d. Budget/resource
    e. Ted
        i. paths to improvement
        ii. Lead from bottom
    f. Don – Tiger Team to document, compare processes between all centers?
        i. Results/outcomes are the key, not the actual processes
        ii. Learn from each other
        iii.
12. Lola
    a. Why started w/ metadata
13. Collect / present metadata stories
    a. who/why/how
    b. 2 minute shorts?
14. Leverage existing work – initially loose, fewer boundaries and structures
    a. Interview process
    b. Video?
15. Jay – no QC system for metadata
    a. Idea for testing: real-world, unscripted scenarios
        i. Identify target communit(ies) for an info asset
        ii. Let 'em know there's a new product and see what they think

iii. OR… see what resources they locate in response to a specific problemspace query
iv. Point 'em at the portal and see if they can find/ acquire/evaluate/ use
v. Collect feedback to eval/improve the metadata record

b. Don – provider review of / feedback on metadata
c. Ted – continual improvement of metadata via incorporation of user input

16. %%Idea – documenting use cases as ISO?
    a. Attempted use is essentially the product name
17. Jay – ECHO experiences applicable to CLASS needs
    a. Successes, failures, approaches
18. Break
    a. Lots of independent discussion, some conceptual, some very specific & technical
19. %%idea – a group consideration/revision of a specific metadata record
    a. Use Jaci Mize as resource
    b. Can do via Webex. MUST have visual component
    c. May be best to do all by Webex, to prevent side discussions that exclude the remote participants
20. 11:10 – Christina – wrap-up / summary
21. Review/ Summary slides
    a. NOAA Brown Bag seminars – ESIP (Carol Meyer, Chris Lenhart)
    b. Cooperation projects / tiger teams / working groups
        i. NCA Metadata Tiger Team – NCDC Metadata Working Group (Waple, Privette, Tilmes)
        ii. CLASS/GOES-R/NASA-ECHO Tiger Team – Granular metadata (Morris, Beaty, Pilone, Mitchell, Baynes, Lighty, Neal, Smith, Ritchey, Lief)
        iii. GCMD (pre-release kick-the-tires)
            1. GOSIC
            2. Climate Portal (climate.gov)
            3. Keywords
        iv. EDMC Wiki (Setup: Jeff DLB, Ted; Population: Christina, Jeff A)
            1. Workshop III presentations
            2. List of tools
            3. Best practices
                a. DataOne examples, etc
                b. Jaci - metadata listserv; international
            4. Discussion forums
            5. Mailing list (this is separate from wiki monitor list, right?)
        v. Videos (or audio) of metadata creation (or use) stories – Lola
        vi. NOAA Data Centers – data/metadata management TT – Process comparison and convergence (Tim O, Jeff DLB, …)
        vii. Metadata User Feedback Tiger Team
22. 11:33 - Tim – closing
    a. Enterprise solutions for data centers within NESDIS

b. Just briefed Mary K on consolidation efforts, plans